



<https://revistas.unan.edu.ni/index.php/Cientifica>
DOI: <https://doi.org/10.5377/esteli.v13i49.17890>

Dimensiones lentamente cambiantes en data warehouses: un análisis a través de la revisión bibliográfica

Slowly changing dimensions in data warehouses: an analysis through literature review

Erick Noel Lanzas Martínez

Universidad Nacional Autónoma de Nicaragua, Managua. UNAN-Managua, Nicaragua

<https://orcid.org/0000-0003-2557-2833>

elanzas@unan.edu.ni

RECIBIDO

15/09/2023

ACEPTADO

16/04/2024

RESUMEN

Este artículo de revisión documental examina el impacto y los desafíos planteados por las dimensiones lentamente cambiantes (SCD) en el contexto de los Data Warehouses. Destaca cómo las SCD presentan desafíos únicos para la gestión de datos y la toma de decisiones empresariales, con un enfoque en la necesidad de mantener datos históricos actualizados y coherentes. Además, se han analizado diversas estrategias de modelado utilizadas por diversos autores para sortear estos obstáculos, incluida la propuesta de Kimball y Ross (2013), así como enfoques como las Temporary Data Warehouses (TDW) y el énfasis en "Extract, Transform, and Load" (ETL). Cada estrategia se adapta a diferentes necesidades de negocio y datos, ofreciendo soluciones valiosas para abordar eficazmente el desafío de las SCD y mejorar la calidad de los datos en los Data Warehouses. Este estudio proporciona ideas valiosas para futuras investigaciones y prácticas, subrayando cómo una gestión estratégica de las SCD puede potenciar la toma de decisiones informadas en la era de la analítica empresarial.

ABSTRACT

This literature review article examines the impact and challenges posed by slowly changing dimensions (SCDs) in the context of Data Warehouses. It highlights how SCDs present unique challenges for data management and business decision making, with a focus on the need to maintain up-to-date and consistent historical data. In addition, a variety of modeling strategies used by various authors to circumvent these obstacles have been discussed, including the Kimball and Ross (2013) approach, as well as approaches such as Temporary Data Warehouses (TDW) and an emphasis on "Extract, Transform, and Load" (ETL). Each strategy is tailored to different business and data needs, offering valuable solutions to effectively address the SCD challenge and improve data quality in Data Warehouses. This study provides valuable insights for future research and practice, highlighting how strategic SCD management can empower informed decision making in the era of business analytics.

PALABRAS CLAVE

Data Warehouses;
Dimensiones Lentamente
Cambiantes (SCD);
Estrategias de Modelado;
Gestión de Datos; Toma de
Decisiones.

KEYWORDS

Data Warehouses; Slowly
Changing Dimensions (SCD);
Modeling Strategies; Data
Management; Decision
Making.



INTRODUCCIÓN

En el mundo empresarial contemporáneo, la toma de decisiones informadas y estratégicas se ha convertido en un elemento crítico para el éxito y la competitividad de las organizaciones. Para lograr este objetivo, las empresas confían en la gestión eficaz de sus datos, y los Data Warehouses desempeñan un papel fundamental como repositorios centrales que almacenan y gestionan información crítica, tal como lo expone Inmon (2005) citado por (Alsahfi, 2020).

Uno de los desafíos más intrigantes y a la vez complejos que enfrentan las organizaciones en el mantenimiento de DataWarehouses es la gestión de dimensiones que cambian lentamente propuesta por Kimball (2008). Estas dimensiones, que representan atributos de datos con cambios graduales a lo largo del tiempo, presentan desafíos únicos para la administración de datos y la toma de decisiones empresariales. En este artículo de revisión bibliográfica, se explora en profundidad el impacto de las dimensiones lentamente cambiantes en el contexto de los Data Warehouses.

Las dimensiones lentamente cambiantes (SCD, por sus siglas en inglés) presentan un dilema interesante: ¿cómo se pueden mantener actualizados y coherentes los datos históricos en un entorno de DataWarehouse a medida que evolucionan las dimensiones? A lo largo de estas páginas, se examinan las implicaciones y los desafíos asociados con SCD, así como las estrategias y mejores prácticas propuestas por expertos en el campo.

Las SCD también simbolizan una de las paradojas más significativas en la gestión moderna de datos: la evolución de los datos a través del tiempo. Esta evolución no es siempre rápida ni disruptiva; a menudo es gradual, reflejando cambios lentos pero constantes en variables como la demografía del cliente, la expansión del producto o las variaciones del mercado, tal como señala Kimball (2008). Las SCD, por lo tanto, representan un aspecto crítico de cómo los datos históricos pueden ser gestionados y utilizados para la toma de decisiones en un entorno empresarial dinámico.

El desarrollo de este artículo se justifica en la necesidad de abordar el desafío que representan las SCD en la gestión de Data Warehouses, destacando su importancia para la toma de decisiones empresariales eficaces y la integridad de los datos. Con los avances tecnológicos y la creciente acumulación de datos, es imperativo comprender cómo gestionar eficientemente estas dimensiones cambiantes para mantener la ventaja competitiva y operativa de las organizaciones. Este estudio pretende aportar claridad y orientación actualizada sobre el manejo de las SCD, ofreciendo así un recurso valioso tanto para académicos como para profesionales del campo.

Los antecedentes de este campo de estudio se remontan a las primeras conceptualizaciones de los Data Warehouses, donde inicialmente se percibía que las dimensiones eran relativamente estáticas, como afirman Golfarelli y Rizzi (2009). Sin embargo, con el tiempo, la necesidad de abordar cómo las dimensiones en un Data Warehouse cambian se hizo evidente, afirma Kimball (2008). Este cambio de paradigma se ha reflejado en la literatura sobre el tema, que se verá analizada en las siguientes páginas.

Este artículo tiene como objetivo proporcionar una visión completa y actualizada de la gestión de Dimensiones Lentamente Cambiantes en Data Warehouses. Se busca proporcionar un análisis del estado actual del conocimiento sobre las SCD, incluyendo los desafíos que presentan y las estrategias para manejarlos, además de identificar áreas de

investigación futuras y ofrecer recomendaciones prácticas para profesionales en el campo. Este enfoque asegura que el artículo sea de valor tanto para la comunidad académica como para los profesionales en el ámbito de la gestión de datos.

MATERIALES Y MÉTODOS

El presente artículo, tiene como base una metodología de orden documental, así mismo se determinó la utilización de un enfoque metodológico centrado en el análisis y síntesis de información y datos existentes sobre un tema específico, a su vez se definió la utilización del método inductivo y deductivo.

Este tipo de investigación se justifica en el contexto del estudio de las Dimensiones Lentamente Cambiantes (SCD) en Data Warehouses, debido a la naturaleza teórica y la amplitud del tema. El análisis documental permite una revisión exhaustiva de la literatura existente, facilitando la comprensión integral de las SCD, sus desafíos y estrategias, según Hernández, Fernández y Baptista (2014). Este enfoque es idóneo para sintetizar y analizar críticamente las diversas perspectivas y contribuciones de expertos en el campo, lo que resulta esencial para abordar un tema complejo y multifacético como el de las SCD.

El tipo de investigación se basa en el análisis documental a través de la consulta de documentos dentro del ámbito de la informática y la gestión empresarial, concentrándose específicamente en las SCD en Data Warehouses. Este estudio aborda tanto la evolución técnica en la gestión de bases de datos como su impacto en la toma de decisiones organizacionales.

Durante la revisión analítica, Hernández, Fernández y Baptista (2014), señalan que el alcance del estudio puede ser descriptivo, exploratorio, correlacional o explicativo. El alcance de este artículo es de carácter exploratorio y descriptivo. Se explora el fenómeno de las SCD en Data Warehouses, describiendo las diferentes aproximaciones y estrategias identificadas en la literatura científica y técnica existente.

Coincidiendo otra vez con Hernández, Fernández y Baptista (2014), el diseño del estudio es no experimental y transversal. Se realiza una revisión en un momento determinado de la literatura existente sin manipulación de variables, enfocándose en describir y analizar el estado actual del conocimiento en el campo de estudio.

El corte del estudio es transversal, pues se consideran estudios y publicaciones en un rango temporal o periodo específico, centrándose en la literatura publicada principalmente en los últimos cinco años, aunque también incluye trabajos fundamentales anteriores para establecer un marco histórico y conceptual.

El enfoque de la investigación es cualitativo, de tipo de teoría fundamentada, priorizando el análisis e interpretación de los contenidos teóricos y conceptuales sobre las SCD en Data Warehouses. Se busca una comprensión profunda de las temáticas tratadas, más que la medición cuantitativa de variables.

El universo de este estudio incluye toda la gama de literatura académica disponible sobre Data Warehouses y SCD. La muestra se seleccionó de manera intencionada para incluir fuentes que proporcionan una visión integral y representativa del tema, centrándose en trabajos de relevancia y actualidad, lo que califica a este artículo como un muestreo no probabilístico decisional.

Materiales

Para la realización del presente artículo fue necesario tener acceso a internet, específicamente al portal de la biblioteca de la UNAN-Managua, que cuenta con acceso remoto a bases de datos académicas y científicas por suscripción, tales como EBSCO, PROQUEST, eLibro, el repositorio institucional de la UNAN-Managua. Otras herramientas utilizadas fueron Microsoft Word, usado para la redacción del documento, y Microsoft Excel, utilizado para facilitar la tabulación de la información recolectada.

Método

El método utilizado para el diseño del presente artículo se basó en el documento de Guirao-Goris et al. (2008) "El artículo de revisión". Es una metodología sistemática y estructurada para identificar, evaluar y sintetizar la investigación relevante sobre el tema. Este método permite no solo la recopilación de información significativa sino también su análisis crítico, asegurando que la revisión sea completa, rigurosa y objetiva, lo que lo convierte en un método válido para lograr los objetivos de este artículo.

El método incluye tres etapas principales, que son a su vez las fases de la investigación.

1. La primera etapa implica precisar la estrategia de búsqueda bibliográfica, definida por la identificación de fuentes de información primaria, secundaria y terciaria, así como las herramientas específicas necesarias para acceder a estas fuentes.

Para la ubicación de los documentos bibliográficos se realizaron búsquedas en la base de datos disponibles a las que se tiene acceso remoto a través de la biblioteca de la UNAN-Managua (biblio.unan.edu.ni). Se consultaron principalmente las bases de datos EBSCO, PROQUES, eLIBRO y RIUMA, el repositorio institucional de la UNAN-Managua utilizando los términos: SCD, Slowly Changing Dimensions, DataWarehouse, Data Warehousing, SCD Data Warehousing Consistency. Además, se realizaron búsquedas complementarias en el buscador "Google Académico" con los mismos términos, priorizando aquellos documentos publicados a partir de 2019 y que fuesen de acceso libre, para garantizar la actualizad y accesibilidad de la información.

2. La segunda etapa se enfoca en la curación del contenido. Esto implica establecer criterios de selección para identificar la bibliografía más relevante al tema de estudio. Esta etapa es crucial para asegurar que solo se incluya información pertinente y de alta calidad. Por tanto, se seleccionaron aquellos documentos que proporcionaban información sobre los aspectos generales sobre DataWarehouse, Dimensiones Lentamente Cambiantes y la problemática que representan, además de incluir trabajos que proponen soluciones, tanto tradicionales, como innovadoras, a la problemática expuesta.

El proceso fue riguroso y metódico, buscando asegurar una cobertura amplia de perspectivas y enfoques sobre el tema.

3. La tercera etapa involucra el análisis de la variabilidad, fiabilidad y validez de las fuentes seleccionadas, garantizando así la credibilidad y la relevancia académica de la información incluida en la revisión.

Respecto a la relevancia y actualidad de las fuentes, se seleccionaron aquellos documentos que tuvieran menos de 5 años de antigüedad, excepto aquellos que

contienen los primeros esbozos de investigación sobre el tema, llamados también fuentes primarias. Estas fuentes primarias son fundamentales para comprender los antecedentes del tema, e independientemente de su fecha de publicación, son relevantes para obtener un marco de referencia completo sobre el tema.

La credibilidad de las fuentes fue una prioridad, seleccionando aquellas fuentes bibliográficas publicadas en repositorios de tesis doctorales, ensayos y artículos científicos, debidamente mediados, revisados y validados por pares y expertos destacados en la temática.

Finalmente se tabularon los resultados en Microsoft Excel, y se procedió al análisis y redacción del documento, usando Microsoft Word 365, optimizado con complementos para la gestión de recursos bibliográficos, específicamente Mendeley Reference Manager.

RESULTADOS Y DISCUSIÓN

Antes de definir la complejidad de las dimensiones lentamente cambiantes, es crucial entender la base sobre la cual se construyen: las dimensiones en un Data Warehouse. Como Kimball y Ross (2013) señalan, estas dimensiones son atributos o características que describen los datos almacenados en las tablas de hechos.

Estas dimensiones proporcionan contexto y permiten organizar y analizar los datos de manera significativa. Es decir, son más que marcadores; son la esencia que da sentido y contexto a los datos brutos. Por lo general, las dimensiones representan entidades del mundo real, como clientes, productos, ubicaciones, fechas y otras abstracciones, y ofrecen un marco para organizar y analizar datos de una manera significativa y aplicable.

Las dimensiones en un Data Warehouse son importantes porque trascienden la mera funcionalidad de organización de los datos; permiten realizar análisis multidimensionales y facilitan la navegación y el filtrado de los datos. Estos análisis son fundamentales en la toma de decisiones empresariales, permitiendo a las empresas navegar y filtrar complejos conjuntos de datos para extraer información valiosa. Además, las dimensiones se utilizan para establecer relaciones con las tablas de hechos, lo que permite realizar agregaciones y cálculos basados en diferentes combinaciones de dimensiones.

Es aquí donde surge el desafío expuesto por Kimball (2008); las Dimensiones Lentamente Cambiantes (Slowly Changing Dimensions (SCD) en inglés) se refieren a las dimensiones en un Data Warehouse que experimentan cambios en sus descripciones a lo largo del tiempo. Estos cambios pueden ser correcciones de errores o cambios reales en la información de una entidad dimensional, como un cliente o un producto. El manejo de estas dimensiones cambiantes no es trivial; es decir, plantea preguntas sobre cómo mantener la integridad y la relevancia de los datos a lo largo del tiempo.

Este planteamiento conlleva entonces a un análisis más profundo de las SCD. Más allá de su definición técnica, se exploran los desafíos intrínsecos que plantean en el contexto de un Data Warehouse dinámico y las estrategias ingeniosas que se han desarrollado para abordar estas complejidades. Es importante destacar que, en este entorno cambiante, las SCD no solo representan un desafío técnico, sino que también encarnan una oportunidad para refinar y perfeccionar la toma de decisiones basada en datos en la era moderna de la analítica empresarial.

El problema de las SCD

El problema de las SCD es cómo manejar los cambios lentos en los atributos de una dimensión en un almacén de datos. Estos cambios pueden incluir actualizaciones, inserciones o eliminaciones de valores en los atributos de una dimensión a lo largo del tiempo. El desafío radica en preservar el historial de cambios en los datos de la dimensión y proporcionar diferentes perspectivas de los datos para informes y análisis (Kimball y Ross, 2013).

Por ejemplo, se supone que se tiene una dimensión de clientes en la cual se almacenan atributos como nombre, dirección y número de teléfono. Si un cliente cambia su dirección, es importante mantener un registro de la dirección anterior para poder analizar datos históricos. El dilema estaría en la necesidad de mantener este registro histórico de direcciones anteriores para análisis retrospectivos, frente a la exigencia de disponer de información actualizada para informes y análisis en tiempo real. Este equilibrio entre preservar el pasado y reflejar el presente es crucial para mantener la integridad y relevancia de los datos en un Data Warehouse. Otro de los ejemplos más utilizados para representar este problema es la evolución de los precios que un producto puede tener a través del tiempo.

El otro problema que han planteado las SCD es el impacto en el rendimiento de las consultas, especialmente en la representación de datos orientados al estado, como señalan Faisal y Sarwar, (2014). Los datos orientados al estado son registros que permanecen válidos durante un período de tiempo, mientras que los datos orientados a eventos son registros que son válidos en un momento particular.

Aquí, el problema se complica: cómo manejar eficientemente estos tipos de datos en un Data Warehouse para que el rendimiento de las consultas no se vea comprometido. La frecuencia de muestreo y la cantidad de datos históricos almacenados pueden influir significativamente en el rendimiento, llevando a una posible redundancia en datos de alta frecuencia o a la pérdida de detalles en datos de baja frecuencia.

Desde una perspectiva analítica, el problema de las SCD no es solo un desafío técnico, sino también una cuestión de estrategia de gestión de datos. La decisión sobre cómo y cuándo actualizar o mantener los datos históricos refleja una elección fundamental sobre el enfoque de una organización hacia su inteligencia empresarial y análisis de datos. La adecuada gestión de las SCD, por lo tanto, se convierte en una herramienta crítica para garantizar que los Data Warehouses sigan siendo recursos valiosos y confiables para la toma de decisiones empresariales informadas.

Calidad y coherencia de los datos en SCD

Tomando como referencia los dos escenarios anteriormente expuestos en la problemática, para el primer caso, significa que las filas de la tabla de hechos deben representar las direcciones antiguas y nuevas para los registros insertados antes y después de la modificación, respectivamente. Por lo tanto, las SCD deben mantener la consistencia y la corrección de los datos como refieren Faisal y Sarwar (2014). La gestión efectiva de las SCD implica no solo reconocer la evolución de los datos sino también adaptar las prácticas de almacenamiento y consulta para acomodar estos cambios sin comprometer la calidad de los datos.

Para el segundo caso, al representar datos orientados al estado utilizando una representación de series temporales, puede perderse información en casos donde la frecuencia de muestreo es baja, o puede producirse redundancia en casos donde la frecuencia de muestreo es alta, lo que afecta significativamente el rendimiento de las consultas. Estos desafíos subrayan la necesidad de un diseño de base de datos y estrategias de consulta que puedan adaptarse a la naturaleza dinámica de las SCD. La implementación de esquemas estrella, tanto regulares con claves sustitutas como temporales, propuesta por Rahayu et al. (2013), ofrece una solución potencial. Sin embargo, estos enfoques pueden aumentar la complejidad de las consultas y las operaciones de unión temporal, lo que plantea preguntas sobre su viabilidad en entornos de Data Warehouse de gran escala.

Impacto en la toma de decisiones

Phungtua-Eng y Chittayasothorn (2022) aseguran que la presencia de SCD puede llevar a problemas analíticos y a imprecisiones en la toma de decisiones. Además, el impacto se destaca en el contexto de la integración de información para actividades de análisis y Business Intelligence (inteligencia de negocios o inteligencia empresarial), donde la precisión y actualidad de los datos son fundamentales.

Por dar un ejemplo contextualizado en la realidad de muchos países latinoamericanos es el de la migración. Supongamos que un almacén de datos recopila datos demográficos, como la edad, el género, la ocupación y la ubicación de residencia de las personas de un determinado país.

En este escenario, se trata de mantener actualizados los datos de residencia de las personas que se han migrado de un país a otro. Algunas personas pueden haber migrado hace años, pero sus datos de residencia en el almacén de datos aún reflejan su ubicación anterior.

La relevancia de mantener actualizados estos datos es crítica. El impacto en la toma de decisiones se manifiesta cuando el gobierno utiliza estos datos para asignar recursos, planificar servicios públicos o tomar decisiones políticas. Por ejemplo, si el gobierno se basa en datos desactualizados y asigna recursos a una región donde ya no reside una gran parte de la población, podría llevar a una distribución ineficiente de recursos y a decisiones erróneas, basadas en una comprensión obsoleta de la distribución poblacional.

En este contexto, el desafío para los Data Warehouses y, por extensión, para la inteligencia empresarial, radica no solo en capturar los cambios en los datos, sino también en interpretarlos y aplicarlos de manera que reflejen con precisión la realidad actual. Esto implica una dinámica de actualización continua y un análisis sofisticado para discernir cuándo y cómo los cambios en las dimensiones afectan las decisiones estratégicas. De allí que el impacto de las SCD influye directamente en la calidad y efectividad de las decisiones basadas en datos.

Estrategias de modelado

Tras la curación de contenido de los artículos analizados, se procedió con la clasificación de la literatura, ordenando las referencias de las más recientes a las más clásicas. La tabla número 1 describe de forma breve el problema abordado por los autores y las estrategias de modelado utilizadas afrontar el problema de las SCD.

Tabla 1: Problemas/Casos de estudio y estrategias de modelado utilizados por los autores.

Autor(es)	Problema abordado / Caso de estudio	Estrategias utilizadas para el modelado
Liu et al. (2021)	Abandono de los datos existentes debido a su mala calidad	<ul style="list-style-type: none"> • Desarrollo de un marco propio que permite parametrizar los modelos más populares para obtener mejores resultados.
Singh y Pandey (2021)	Gestión de campañas de marketing en instituciones financieras y las dificultades asociadas con el manejo de datos en un data mart dependiente	<ul style="list-style-type: none"> • Creación de una tabla de datos dependiente (DCHT). • Aplicación de SCD Tipo 2 del modelo Kimball
Gupta y Sahayadhas (2020)	Diseño eficiente de almacenes de datos para mejorar los sistemas de soporte de decisiones en empresas y organizaciones	<ul style="list-style-type: none"> • Modelado ágil de información, la transformación y carga de datos (ETL), la optimización del rendimiento y la seguridad de los datos
Vaisman y Zimányi (2019)	Construcción de almacenes de datos de movilidad que permitan realizar consultas analíticas en línea (OLAP) combinadas con consultas de objetos móviles.	<ul style="list-style-type: none"> • Extensión del modelo conceptual MultiDim • Uso de tipos de datos temporales • Integración de MobilityDB
Yang et al. (2019)	Identificar, analizar y comparar diferencias y tendencias de DWHAs desde las perspectivas de componentes y arquitectónicas.	<ul style="list-style-type: none"> • Se compararon 9 modelos DWHAs representativos en una "imagen general" que muestra directamente sus similitudes y diferencias.
Rahayu et al. (2013)	Implementación de (SCD) para apoyar la formación de dimensiones dinámicas en un Data Warehouse.	<ul style="list-style-type: none"> • Aplicación de Slowly Changing Dimensions (SCD) type 2 • Implementación de la metodología Nine-Step Methodology de Kimball • Creación de tablas de hechos y dimensiones: • Transformación de datos para la integración de datos transaccionales y maestros • Implementación de Star Schema en la base de datos OLAP

<p>Kimball y Ross (2013)</p>	<p>Manejar los cambios lentos en las dimensiones en un almacén de datos</p>	<ul style="list-style-type: none"> • Tipo 1 (sobrescribir) • Tipo 2 (agregar nueva fila) • Tipo 3 (agregar nuevo atributo) • Tipo 4 (agregar mini-dimensión) • Tipo 5 (agregar mini-dimensión y tipo 1 adicional) • Tipo 6 (agregar atributos tipo 1 a una dimensión tipo 2)
<p>Golfarelli y Rizzi (2011)</p>	<p>Gestión de las dimensiones que cambian lentamente en un almacén de datos.</p>	<ul style="list-style-type: none"> • Se propone el modelo de datos espaciotemporales • Se proponen la evolución del esquema y la transferencia de cambios de esquema relacionados con el esquema de vista antiguo en vistas materializadas, usando Temporal Data Warehouse (TDW)
<p>Singh (2006).</p>	<p>Problema de la consistencia en las SCD.</p>	<ul style="list-style-type: none"> • Implementación basada en análisis de la herramienta comercial Informatica para resolver el problema de consistencia de datos en almacén de datos en presencia de cambios lentos en las dimensiones. • Implementación de “Extract, Transform, and Load” (ETL)
<p>Moody y Kortink (2000).</p>	<p>Desarrollo de modelos dimensionales a partir de modelos de entidad-relación tradicionales para el diseño de almacenes de datos y Data Marts basados en modelos de datos empresariales.</p>	<ul style="list-style-type: none"> • Clasificación de entidades • Identificación de jerarquías • Diseño de Data Marts

De la tabla anterior, se destacan las estrategias de modelado de Kimball y Ross (2013) donde proponen diferentes estrategias de modelado para abordar el problema de las SCD, como el tipo 1 (sobrescribir), tipo 2 (agregar nueva fila), tipo 3 (agregar nuevo atributo), tipo 4 (agregar mini-dimensión), tipo 5 (agregar mini-dimensión y tipo 1 adicional) y tipo 6 (agregar atributos tipo 1 a una dimensión tipo 2). Cada estrategia tiene sus ventajas y desventajas, y la elección de la estrategia depende de los requisitos específicos del negocio y de los datos. Por ello, es crucial considerar que la elección de la estrategia adecuada debe basarse no solo en los requisitos técnicos sino también en los objetivos estratégicos de la organización, lo que a veces puede requerir un equilibrio delicado entre precisión de los datos y complejidad del sistema.

Otros autores analizados, utilizaron estas herramientas como insumo para crear sus propias estrategias de modelado, tal es el caso de Rahayu et al. (2013), Yang et.al (2019), Singh y Pandey (2021) y Liu et al. (2021). Resulta curioso que en los cuatro trabajos se haya utilizado la estrategia del Tipo 2 Propuesta por Kimball y Ross (2013). Este tipo de

adaptaciones son testimonio de la flexibilidad y la robustez del marco original de Kimball y Ross (2013), aunque las adaptaciones deben ser cuidadosamente planificadas para evitar complicaciones innecesarias.

De aquí surge una nueva interrogante: ¿Qué pasa cuando se agregan demasiadas filas a la dimensión? la gestión efectiva de este problema es fundamental, ya que una acumulación excesiva puede degradar el rendimiento del sistema y la agilidad en la toma de decisiones. Por ello, Liu et al. (2021) se aventuran a diseñar un marco propio basado en encontrar una estrategia de control óptimo fundamentado en tuplas matemáticas que contienen las relaciones entre las tuplas y la importancia/calidad de la información contenida. Por su parte, Singh y Pandey (2021) manejan este problema creando una tabla de datos dependiente y apuntando las relaciones de las filas agregadas a esta nueva tabla.

Rahayu et al. (2013), manejaron el problema implementando Star Schema en la base de datos OLAP, de esta forma se gestionó la transformación de datos para la integración de datos transaccionales y maestros. Adicionalmente, se crearon las tablas de hechos y dimensiones para referenciar a las nuevas filas agregadas. De hecho, la estrategia de Singh y Pandey (2021) es una mejora a esta última estrategia. Vale la pena destacar esta estrategia por su capacidad para mantener la organización y la claridad en el Data Warehouse, aunque de forma personal, el autor del presente artículo destaca que puede ser desafiante implementarla en sistemas ya complejos y establecidos.

Otras de las estrategias utilizadas es la implementación de las Temporary Data Warehouses (TDW), que según Golfarelli et al. (2009), sirven de bases de datos temporales y se combinan con un Data Warehouse que pueden representar el cambio de los datos a través del tiempo. Esta estrategia fue la utilizada en los trabajos de Vaisman y Zimányi (2019) y Golfarelli y Rizzi (2011), aunque ambos desde casos de uso totalmente distintos.

En el caso de Vaisman y Zimányi (2019) tenían la necesidad de acceder datos de movilidad que permitan realizar consultas analíticas en línea (OLAP) combinadas con consultas de objetos móviles, ya que se analizaba un caso de uso para gestiones de transporte. Por otro lado, Golfarelli y Rizzi (2011) se centraron más en la revisión y documentación previa del problema planteados por Inmon (2005) y Kimball (2008) proponiendo la aplicación de las TWD propuestas por Golfarelli et al (2009).

El autor de este artículo valora positivamente la TDW por su potencial para ofrecer una representación más dinámica y temporal de los datos, sugiriendo que esta podría ser una dirección valiosa para futuras investigaciones y aplicaciones prácticas.

Una tercera técnica utilizada, es la de implementar “Extract, Transform, and Load” (ETL), utilizado por Gupta y Sahayadhas (2020) y Singh (2006). Este proceso es fundamental en la gestión de datos, especialmente en el contexto de la integración y preparación de datos para su almacenamiento y análisis en un almacén de datos. Se basa en la gestión de la información previa a su almacenamiento en el Data Warehouse. Estrictamente hablando, no es necesariamente una nueva técnica, sino que ETL es parte del proceso de creación de los Data Warehouses, pero estos autores afirman que una buena ejecución de este proceso contribuye enormemente a la solución del problema.

Aquí es necesario subrayar la importancia de un proceso de ETL bien diseñado y ejecutado, destacando que es un pilar fundamental para el éxito de cualquier estrategia de Data Warehouse, especialmente en entornos donde las SCD son una preocupación significativa.

Finalmente, trabajos como el de Moody y Kortink (2000) sirvieron como insumo a Kimball (2008) para la propuesta de sus estrategias. Este trabajo toma como referencia el ya conocido modelo relacional de Codd (1970) para el diseño de Data Marts que pudiesen gestionar las dimensiones lentamente cambiantes.

La tabla número 2 resume las tres principales estrategias de modelado utilizadas por los distintos autores:

Tabla 2: Estrategias de modelado utilizadas por los autores

Autores	Estrategias de Modelado
Singh y Pandey (2021) Liu et al. (2021) Yang et.al (2019) Rahayu et al. (2013)	Tipos de Kimball & Ross (2013), con ajustes/modificaciones.
Vaisman y Zimányi (2019) Golfarelli y Rizzi (2011)	Temporary Data Warehouses (TDW)
Gupta y Sahayadhas (2020) Singh (2006)	Énfasis en “Extract, Transform, and Load” (ETL)

Como se hace evidente en la tabla número 2, las propuestas de Kimball y Ross (2013) son ampliamente utilizadas, aunque frecuentemente adaptadas para ajustarse a las necesidades y contextos específicos de cada caso. Estas adaptaciones reflejan la flexibilidad y la relevancia continua de las propuestas de Kimball y Ross en la gestión de SCD, permitiendo a los usuarios mantener la integridad de los datos mientras se adaptan a cambios.

Por otro lado, las estrategias TDW y ETL han demostrado ser efectivas. Los TDW, en particular, ofrecen un enfoque interesante para manejar las SCD, proporcionando un marco que puede soportar mejor la volatilidad de los datos y su representación temporal. Esta estrategia se destaca por su capacidad para reflejar de manera más dinámica y precisa el cambio en los datos a lo largo del tiempo.

A criterio del autor de este artículo, la adopción de TDW representa una prometedora línea de investigación para futuros estudios. Explorar más a fondo cómo las TDW pueden integrarse de manera efectiva en los sistemas de Data Warehouses existentes podría abrir nuevas vías para mejorar el manejo de datos dinámicos y complejos, potenciando así la capacidad de las organizaciones para realizar análisis más precisos y fundamentados.

En cuanto al ETL, su papel es fundamental en la preparación de los datos para su análisis. Aunque no es una técnica nueva, su implementación efectiva es crucial para el éxito de cualquier estrategia de Data Warehouse, especialmente en escenarios donde las SCD son prevalentes. La continua evolución de las técnicas de ETL para manejar mejor las SCD podría resultar en mejoras significativas en la eficiencia y efectividad de los Data Warehouses.

Las tres principales estrategias aplicadas por los diversos autores no solo ofrecen soluciones técnicas, sino que también plantean importantes consideraciones estratégicas que pueden influir profundamente en la inteligencia empresarial y la toma de decisiones basada en datos.

CONCLUSIONES

El presente artículo de revisión ha explorado el impacto de las dimensiones lentamente cambiantes (SCD) en el contexto de los Data Warehouses. Se ha destacado cómo las SCD presentan desafíos únicos para la gestión de datos y la toma de decisiones empresariales. Estos desafíos incluyen la necesidad de mantener datos históricos actualizados y coherentes, así como el impacto en la calidad de los datos y la toma de decisiones.

Además, se han analizado y discutido diversas estrategias de modelado utilizadas por expertos en el campo para abordar el problema de las SCD, destacando la propuesta de Kimball y Ross (2013) y otras como las Temporary Data Warehouses (TDW) y el énfasis en “Extract, Transform, and Load” (ETL). Cada estrategia presenta ventajas y desafíos, y su elección depende de las necesidades específicas de negocio y datos. Las estrategias de modelado discutidas en este artículo brindan herramientas valiosas para abordar eficazmente el problema de las SCD y mejorar la calidad y coherencia de los datos en los Data Warehouses.

Las discusiones en este artículo resaltan que la elección adecuada de la estrategia de modelado no solo responde a necesidades técnicas, sino que también refleja un compromiso estratégico con la mejora continua de la infraestructura de datos. Las TDW, por ejemplo, ofrecen oportunidades para gestionar la volatilidad de los datos de forma más eficaz, lo que puede ser crucial para organizaciones que operan en entornos dinámicos. La revisión sugiere que futuras investigaciones podrían explorar más a fondo la integración de técnicas avanzadas de ETL con TDW para crear soluciones más robustas y escalables que soporten mejor las necesidades de los modernos Data Warehouses. Además, sería provechoso examinar el impacto a largo plazo de la implementación de estas estrategias en la operatividad y la toma de decisiones en diversas industrias, especialmente en aquellas con grandes volúmenes de datos y requisitos de rapidez en la actualización de información.

En última instancia, este artículo no solo sintetiza el estado actual del conocimiento sobre las SCD en Data Warehouses sino que también, ofrece una perspectiva crítica sobre cómo las organizaciones pueden implementar estrategias de modelado efectivas para superar estos desafíos. Por lo tanto, se convierte en un recurso invaluable para profesionales de la gestión de datos e investigadores, proporcionando una base sólida para abordar este desafío crítico en la era de la analítica empresarial.

REFERENCIAS BIBLIOGRÁFICAS

- Alsahfi, T., Almotairi, M., & Elmasri, R. (2020). A survey on trajectory Data Warehouse. *Spatial Information Research*(28), 53-66. doi:<https://doi.org/10.1007/s41324-019-00269-x>
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377-387. Obtenido de <https://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>
- Faisal, S., & Sarwar, M. (2014). Handling slowly changing dimensions in data warehouses. *Journal of Systems and Software*(94), 151-160. doi:<https://doi.org/10.1016/j.jss.2014.03.072>
- Golfarelli, M., & Rizzi, S. (2009). A Survey on Temporal Data Warehousing. *International Journal of Data Warehousing and Mining*, 5(1), 1-17. doi:<https://doi.org/10.4018/jdwm.2009010101>
- Golfarelli, M., & Rizzi, S. (2011). Temporal Data Warehousing: Approaches and

- Techniques. En D. Taniar, & L. Chen, *Integrations of Data Warehousing, Data Mining and Database Technologies: Innovative Approaches* (págs. 1-18). IGI Global. doi:<https://doi.org/10.4018/978-1-60960-537-7.ch001>
- Guirao-Goris, J. A., Olmedo Salas, Á., & Ferrer Ferrandis, E. (2008). El artículo de revisión. *Revista iberoamericana de enfermería comunitaria*, 1(1), 1-25. Obtenido de https://www.researchgate.net/publication/268518751_El_articulo_de_revision
- Gupta, A., & A., S. (2020). A Comprehensive Survey to Design Efficient Data Warehouse for Betterment of Decision Support Systems for Management and Business Corporates. *International Journal of Management (IJM)*, 11(7), 463-471. Obtenido de https://iaeme.com/MasterAdmin/Journal_uploads/IJM/VOLUME_11_ISSUE_7/IJM_11_07_044.pdf
- Gupta, A., & Sahayadhas, A. (2020). Proposed Techniqueto OptimizetheDWandETLQuery for Enhancing data warehouse efficiency. *5th International Conference on Computing, Communication and Security (ICCCS)* (págs. 1-5). Patna: IEEE. Obtenido de <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9276824&isnumber=9276491>
- Hernández, R., Fernández, C., & Baptista, M. d. (2014). ¿En qué consiste la revisión analítica de la literatura? En R. Hernández Sampieri, C. Fernández Collado, & M. d. Baptista Lucio, *Metodología de la investigación* (6° ed.) (pág. 61). México D.F: McGraw Hill Interamericana Editores S.A. de C.V.
- Hernández, R., Fernández, C., & Baptista, P. (2014). Concepción o elección del diseño. En R. Hernández, C. Fernández, & P. Baptista, *Metodología de la investigación* (6° ed.) (pág. 127). México D.F: McGraw Hill Interamericana Editores S.A. de C.V.
- Hernández, R., Fernández, C., & Baptista, P. (2014). Definición del alcance de la investigación que se realizará: exploratorio, descriptivo, correlacional o explicativo. En R. Hernández, C. Fernández, P. Baptista, & S. D. INTERAMERICANA EDITORES (Ed.), *Metodología de la Investigación* (6° ed.) (pág. 90). México D.F: McGraw Hill Interamericana Editores S.A. de C.V.
- Inmon, W. H. (2005). Going from the Data Warehouse to the Operational Environment. En W. H. Inmon, *Building the Data Warehouse* (4ta ed., pág. 117). Nueva Jersey: Wiley.
- Kimball, R. (2008). Slowly Changing Dimensions. *DM Review*, 18(29), 1-9. Obtenido de <https://www.proquest.com/scholarly-journals/slowly-changing-dimensions/docview/214686457/se-2?accountid=171676>
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3ra ed.). Indianapolis: Wiley.
- Liu, Q., Feng, G., Tayi, G., & Tian, J. (Abril de 2021). Managing Data Quality of the Data Warehouse: A Chance-Constrained Programming Approach. *Information Systems Frontiers, Springer*, 23(2), 375-389. doi:<https://doi.org/10.1007/s10796-019-09963-5>
- Moody, D. L., & Kortink, M. A. (2000). From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. *Proceedings of the Second Intl. Workshop on Design and Management of Data Warehouses* (págs. 1-12). Estocolmo: DMDW 2000. Obtenido de <https://ceur-ws.org/Vol-28/paper5.pdf>
- Phungtua-Eng, T., & Chittayasothorn, S. (2022). Information Integration and Multiple Slowly Changing Dimensions Modeling. *ICCMS '22: Proceedings of the 14th International Conference on Computer Modeling and Simulation*, (págs. 214-222). New York: Association for Computing Machinery. doi:<https://doi.org/10.1145/3547578.3547611>
- Rahayu, G. P., & Gunawan, R. (2013). Penerapan Slowly Changing Dimensions untuk Mendukung Pembentukan Dimensi Dinamis pada Data Warehouse (Studi Kasus: Dinas Kependudukan dan Pencatatan Sipil Kabupaten XYZ). *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 33-38. Obtenido de <https://journal.uui.ac.id/Snati/article/view/3062/2822>
- Singh, A., & Pandey, S. P. (2021). Study of Slowly Changing Dimension to Dependent Data Mart to Manage Finance Data. *International Journal of*

- Scientific Research in Computer Science, Engineering and Information Technology, 7(6), 190-195. doi:<https://doi.org/10.32628/CSEIT217657>
- Singh, B. (2006). Implementation of a solution to the slowly changing dimension (SCD) data warehouse consistency problem using extract transform and load (ETL) technology. Northridge: California State University. Obtenido de <https://scholarworks.calstate.edu/downloads/4t64gs61v>
- Singh, R., & Singh, K. (Mayo de 2010). A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. *IJCSI International Journal of Computer Science*, 7(2), 41-50. Obtenido de <https://courseware.cutm.ac.in/wp-content/uploads/2020/06/Descriptive-Classification.pdf>
- Vaisman, A., & Zimányi, E. (2019). Mobility Data Warehouses. *International Journal of Geo-Information*, 8(4), 170. doi:<https://doi.org/10.3390/ijgi8040170>
- Yang, Q., Ge, M., & Helfert, M. (2019). Analysis of Data Warehouse Architectures: Modeling and Classification. *Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 2: ICEIS* (págs. 604-611). Creta: SciTePress. doi: <https://doi.org/10.5220/0007728006040611>